AI DIGEST

A mathematician announced a solution to a 78-year-old problem - a complex structure on S⁶, done with Claude. 370K views overnight

AI digest, Monday 24.08 - the day the question "which model is best" finally turned into "which model for which task": a mathematician closed a 78-year-old problem with Claude, and payment data showed that almost nobody buys the most expensive model.

Monday, 24 August 2026

IN THIS ISSUE

TOPIC 1

A mathematician announced a solution to a 78-year-old problem - a complex structure on S⁶, done with Claude. 370K views overnight

Post by Levent Alpöge (@__alpoge__) (24.08, 00:31 Kyiv time) - **370.7K views, 1.7K likes** in under seven hours. Verbatim: "Please welcome to the world a beautiful new geometric object, to do with a problem i've always loved. **claude really contains multitudes:D** Does S^6 admit a complex structure? / **Yup**".

What the problem is. Whether the six-dimensional sphere admits a complex structure has been open **since 1948**. The last known attempt to prove that it does not was by the late Michael Atiyah, one of the biggest names in 20th-century geometry; his preprint comes up in the replies under the post.

The important part: the author published a checkable certificate. A PDF at alpo.ge/s6.pdf - **1.1 MB, 362K characters of text**, downloaded and read. It is dense differential geometry: the triangle group D(3,4,∞), a universal family of 2-tori over a stacky P^1, three special fibers (Kodaira logarithmic transformation at the elliptic points, Mumford toric degeneration at the cusp). The final claim, verbatim: "**X has the integral homology of S^6 and is diffeomorphic to S^6**".

The author's motivation in a second post: "i always like to have short certificates that people can use to verify in full with their expertise **and/or with one of today's super amazing models**".

And the antidote straight away, because this is exactly the class of topic that needs one. This is **a claim, not a peer-reviewed result**:

- The work **contradicts a published result** by Campana - Demailly - Peternell. The PDF has a section titled "Why the argument of [CDP20] does not apply": the author is knowingly saying the published paper has a flaw. @proofchecker under the post adds: "Interestingly, **the whole point of CDP20 was to fill gaps in CDP98**".
- The soberest comment, @ccccccmore: "announcing a new geometric object with a chatbot as co-author is a bold way to say **the referees haven't seen it yet**".
- @IbrahimDagher20: "if claude did this ~all, **this may be a first for AI?**". Even the supporters end on a question mark.

The context that makes this serious: **this is the same Alpöge** whose counterexample to the Jacobi conjecture, done with Fable, pulled 803 points on HN in July. He is a mathematician with a PhD under Manjul Bhargava, a Fields medallist.

Why it matters. The practical takeaway is about the **format of verification**. Alpöge does what 99% of "AI solved something" posts skip: a short self-contained certificate you can hand to an expert OR to another model. That is exactly the discipline this digest holds to: **an artifact you can check outweighs a claim**. Tag: [promising], not [proven], pending review.

TOPIC 2

Payment data from 70K companies: Opus 5 takes 3.5% of Anthropic spend, Opus 4.8 takes 28%. The best model loses to its own predecessor sixfold

Simon Willison's write-up (23.08, 20:24) of an FT piece, plus the Ramp index - an estimate built on **billing data from 70,000 companies** paying with Ramp corporate cards. Not on surveys.

Share of spend on Anthropic models, July 2026:

MODEL	SHARE
Opus 4.8	**28.0%**
Sonnet 4.6	8.3%
Fable 5	**8.0%**
Opus 4.6	6.9%
Sonnet 5	3.6%
Opus 5	**3.5%**
Haiku 4.5	1.0%

Willison corrects himself immediately: Opus 5 shipped on **24 July**, so it only caught a week of the July data, and a low share is expected. But **Fable 5 at 8.0% given its price** is a signal, and that is what the FT headline is about.

The money figures from the same piece (FT, citing "people with knowledge of the matter"): Anthropic's annualised revenue **$65B** in July against **$47B** in May; **6,000 customers** paying $100K/year or more; Q3 expected to be profitable. OpenAI is at "over **$40B**, +35% quarter over quarter".

Dedup, and honesty about the source filter. This figure already appeared on **18.08**, and back then it came from a `@AiBreakfast` post, an account **outside the subscriptions** that had been wrong before. It was checked against CNBC and Bloomberg then. Today the FT confirms it. The filter "a doubtful source sends you to check a solid one" worked, and the figure survived three checks.

What could not be verified, stated plainly: the FT piece itself could not be read - **ft.com is blind**: it returns 403 both on the real address and on a made-up one (`00000000-0000-...`), so curl measures nothing there. The figures come **from Willison's summary**, credited to that source. The Ramp index, by contrast, checks out honestly: 200 on the real page, **404 on a made-up one**.

Why it matters. In July the market sat overwhelmingly on **Opus 4.8**. "Everyone is already on the latest model" is a Twitter illusion; the money says otherwise.

TOPIC 3

Breunig turns it into a thesis: "Fable and the end of the free lunch". The Moore's law analogy holds

A post by Drew Breunig (23.08), quoted by Willison separately, and the most useful text of the day.

The thesis, verbatim: "Prior to Fable, it felt silly to waste too much time improving your coding harness or context strategies. **A new model would arrive at the same price (or**

cheaper!) and paper over most of your problems. But then Fable landed... So we started to think about **what work went where**".

The analogy: while Moore's law held, polishing code made little sense, because in 18 months the processor would double its speed by itself (Herb Sutter called this "the free lunch"). Once single-thread performance stalled, people had to think about parallelism and architecture. **The same thing has now happened with models.**

A concrete number: **GLM 5.2 shipped the same week as Fable and costs about 1/9 of it** (and about 1/5 of Opus 5). "Is GLM 1/9th the quality of Fable? Perhaps, for certain classes of tasks. But **for most rote coding it's more than sufficient. Especially when provided with great context**".

The author's own working pattern, usable as is: "I frequently chat with Fable to **interrogate and shape a design**, before handing off a brief to GLM".

And the counterargument he raises against himself: inference prices will keep falling, so won't everything go back to "run it all through the biggest model"? His answer: no, because **the same price drops reach K3 and Qwen too**, and better harnesses make weaker models more usable.

Why it matters. The scaffolding around a model - instructions, memory, tested recipes - is the "harness" that lets a weaker model do good work. The move this suggests: **split tasks by class**. Analysis and writing go to the heavy model. Mechanical work (rewriting a script from an existing template, parsing, reformatting) goes to a cheaper one.

TOPIC 4

Prime Intellect: 153 autonomous runs, 18 models, real harnesses. Fable closed 81.7% of the gap to the human record, Opus 5 closed 53.6%

NanoGPT Speedrun Frontier - **127 points on HN**, thread. The valuable part is that the models were run **through the same harnesses developers actually use**: `claude-code`, `codex`, `grok-cli`, `kimi-code`.

The task: autonomously optimise a nanoGPT training recipe, reaching the target loss of 3.28 in fewer steps, within a time and token budget.

MODEL	HARNESS	GAP CLOSED
Fable 5	claude-code · high	**81.7%** (8.7 days)
Opus 5	claude-code · max	**53.6%** (2.9 days)
Kimi K3	prime-agent · max	52.2%
Opus 4.8	claude-code · max	39.4%
GPT-5.6 Sol	codex · xhigh	35.9%
Sonnet 5	claude-code · max	26.8%
GLM 5.2	pi · high	20.3%

And a hole in the methodology, found on HN right away. A comment from @ninjahawk1: "why was **Fable 5 tested on high while Opus 5 was tested on max**? Seems like quite a few of them aren't on the same effort setting... that might be viewed as an **experimental error**". Fable's first place was measured at a different effort level than Opus in second, so the columns cannot be compared directly.

A second comment, @Farmadupe, goes after the report's prose: phrasing like "a frozen verify.py accepts the claim" reads to him as generated slop - "what does it mean to freeze a python script?".

Why it matters. Take **the measurement itself**: on a real multi-day autonomous task the difference between models is measured **in multiples**. And the same model at a different effort level gives a different result: **effort is a real lever**.

TOPIC 5

Torvalds debugged with an AI and described it exactly as it is: "several times stated flat out that this was impossible and unsolvable"

A quote collected by Willison from an actual kernel commit (drm/xe).

Verbatim: "And this was a debug session from hell, **enormously helped by an AI** doing much of the grunt-work. I'd like to call it my tireless helper, but the AI **several times stated flat out that this was impossible and unsolvable** and that we should just write a report about it. I suspect those things have been **trained by people who may not be quite as stubborn as I am**. But while the AI was ready to give up several times, it did keep adding debug code and analyzing it faithfully **when I pushed**. So credit where credit is due and **I let the AI write the commit message** above".

Why it matters. The most useful note of the day for daily work: **a model gives up before the options are exhausted**. Torvalds got a result only because he kept pushing. Worth holding as a plain rule: when a model says "this is impossible / there is no data", that is a hypothesis,

and it deserves another push. It is the same lesson as "evidence contradicts it, so the hypothesis falls", seen from the side of the person doing the pushing.

TOPIC 6

Qwen 3.8 27B on a single machine reverse-engineered a licence check in 30 minutes

An XDA article (22.08) - **159 points on HN**, thread.

The hardware is named honestly: a Lenovo ThinkStation PGX on NVIDIA GB10 Grace Blackwell, **128 GB of unified memory**, 273 GB/s. Out of the box 15-30 tokens/s; with SGLang + NVFP4 + DFlash2 it reaches **about 50 tokens/s** on code. The harness is Pi, and the model only called standard bash tools. By Artificial Analysis's rating it is the **top open-weights model in the 4-40B class out of 135 models** (intelligence index 52).

The most interesting comment in the thread, @saidinesh5: the future is **large frontier models generating and updating skills for "good enough" local models**. Many tasks do not need that much compute; good documentation and skills are enough.

Why it matters. This is item 3 from the other end, with the same conclusion: value shifts away from model size and towards **instructions and context**. "The frontier model writes the instructions, the local model executes" is the same division of labour, just running on your own hardware.

TOPIC 7

Stripe gave agents a wallet: a CLI that issues single-use payment credentials. 675 stars, MIT

An announcement from @jeff_weinstein (23.08, 32K views), reposted by Patrick Collison. The repo is stripe/link-cli, checked via the API: **675 stars, MIT licence**, last push 20.08.

The description, verbatim: "Let your agents spend on your behalf. **Your payment credentials are never exposed. You approve every purchase.**" The mechanics: the agent gets either a virtual card (works at any checkout, not only Stripe) or a Shared Payment Token for merchants that support Machine Payment Protocols. There is a **local MCP server** and an instruction-set install: `npx skills add stripe/link-cli`.

The limitation is right in the README: "For now, this is **only available to US Link accounts**". Outside the US it is not usable today.

Why it matters. The form factor is worth noting: money for an agent is packaged as **a single-use delegated permission with a human confirming every purchase**. It is the "a dangerous action needs explicit approval" model translated into a payment protocol. By the time it reaches the EU the architecture will already be familiar.

TOPIC 8

Ronacher: AI made the choice of programming language less consequential, and people moved to the "hard" languages

"Fast and Hard Code" by Armin Ronacher (22.08) - **81 points on HN**, thread.

The thesis: "the act of **familiarizing yourself with a language no longer matters** and some of the friction that mattered for humans does not matter for agents. As a result, **LLMs make language choice much less consequential** than it used to be".

The consequence he observes: people started picking languages **by marketing**, and Rust and even Zig are the winners, despite part of the core Zig community being hostile to AI. The examples are specific: Cloudflare's new Artifacts service has a **Git protocol engine in pure Zig compiled to ~100 KB of WebAssembly**; Vercel shipped **fx**, a coding agent written in Zig.

Why it matters. A sober backdrop to item 3: if choosing a language gets cheaper, what gets more expensive is what cannot be automated - **architecture, verification and taste**. It also explains why the feed is suddenly full of Zig.

TOPIC 9

Garry Tan: "systems of record will need to become AI harnesses or face replacement by agents"

A post by @garrytan (24.08, 04:22 Kyiv time - fresh, 2.4K views in nine minutes at collection time). One sentence with nothing unpacked, so it goes in as a thesis: "Prediction: **systems of record will need to become AI harnesses or face replacement by agents**".

Worth including because it is the same word "harness" from items 3, 4 and 6, applied to products: value moves from "where the data sits" to "what an agent can do with it".

TOPIC 10

Mollick on a side effect of bots and on honesty in AI research

Two observations from @emollick over the day, both small, but they fit the theme.

The first (24.08, 02:53, **7.9K views**): "Annoying side effect of all the AI bots on the site is that they are all "**well-read**" and therefore reply cogently - my niche reference tweets that would normally attract a small but interested group **now get LLM replies**". A niche signal on Twitter stops working as an insider filter.

The second, more important, on methodology (23.08, 21K views): publishing AI-impact research run on **old models** is not a crime, but "it requires a **very careful discussion** and has to be **very clear to non-technical readers**". That is exactly the problem daily work with numbers runs into: a figure from an April model, presented without a date, reads as a figure about today.

misc: Mollick on consumer AI - 34K views, "healthcare, government, personal finance, school forms" are areas where people "muddle by, but need help they can't get", which is why consumer AI is underrated · the same author on LLM financial advice - MIT and Stanford research: for most people a model's advice beats their own decisions, but quality depends on **what questions get asked** · @amasad: "a week has 7 days, so 7 releases" 26K views, Replit's week: Free Mode, Conversations, Routines, importing instruction sets from GitHub, black-box pentests · @lennysan on enterprise sales 266K views - "if your win rate is above 35%, your price is too low", 15 stages instead of 5 · @levie: "I'm good to retire now" 240K views · @levelsio on plastic at home - 129K views, furniture and rugs are almost universally synthetic; separately on Greece - **85% of homes have air conditioning against 20% on average across Europe** · @ClementDelangue on NVIDIA AVO 42K views - yesterday's item 1, now from the Hugging Face side · HN's death room - malware in the firmware of Android car head units, 87 points

---